Motion Planning in Understructured Road Environments with Stacked Reservation Grids

Fangyu Wu[1†], Dequan Wang[1], Minjune Hwang[1], Chenhui Hao[1], Jiawei Lu[1], Trevor Darrell[1], and Alexandre Bayen[1]

***Abstract*— Motion planning of autonomous vehicles in understructured road environments is challenging owing to a lack of an efficient and analyzable representation of the contextual information. On one hand, idealistic representations like splines, while efficient and interpretable, are not versatile enough to encode the full complexity of the environment. On the other hand, high-fidelity representations like images, although rich in contents, are computationally expensive to decode and are not readily analyzable. To address this problem, we propose a new data structure named stacked reservation grid (SRG). We show that SRG can encode a vast amount of information from the environment without loss of computational efficiency and analytical tractability. We first define SRG by drawing a comparison with the classical data structure, i.e., the occupancy grid. Next, we describe how to compute SRG offline from a continuous history of occupancy grids. After, we propose a neural network model to estimate SRG from a recent motion history of the surrounding objects. Finally, we show that SRG can be efficiently applied in a linear program to solve motion planning. To test our proposed approach, we build a dataset named the Berkeley Deep Drive Drone (B3D) dataset and describe how to develop a validation procedure using the B3D dataset. The dataset and related codes will be open sourced at https://fywu85.github.io/b3d/.**

I. Introduction

A decade of research and development in autonomous driving has enabled us to build automated vehicles that can operate effectively in a variety of road environments. However, fundamental challenges in perception, planning, and verification are still present, significant enough to prevent us from deploying autonomous vehicles to *all* conditions of the roads [1]. Among those challenges, motion planning, while solved for common driving scenarios, still remains a challenge for navigation in understructured road environments like unsignalized intersections, roundabouts, and overcrowded highways.

A fundamental problem for motion planning in understructured environments is how to encode sufficient information from the environment in a mathematically tractable data structure that imposes minimal time and space complexity on the models that use it. Conventional ideas for such representations include image [2], lattice [3], graph [4], and field [5]. By construction, the choice of representation dictates the structure of the higher-level motion planning algorithms. Image-based representation ties strongly to computer vision and enables direct adaptation of deep neural network models [6]. Lattice-based representation favors techniques such as parameter space search [7]. Graph-based representation enables applications of search algorithms like hybrid A* [8] and RRT* [9] on a graph-based map. Field-based representation allows for a more physics-motivated approach such as artificial potential field [5] and vector field histogram [10].

Conventional approaches are designed in a different contexts and therefore have notable limitations for motion planning in understructured environments. Image-based representations can store rich information from the environment but needs a powerful decoding scheme and is not known to have a way to certificate performance. Lattice-based and graph-based methods are efficient for lower dimensional path planning but (1) becomes exponentially more expensive in both storage and computation in higher dimensional space containing control various parameters and time, (2) and suffers from discretization errors. Field-based methods are fast heuristic path planner good for simple navigation tasks but produces suboptimal or even infeasible solutions for complex field topology.

In light of the past work, we propose a new representation method named *stacked reservation grid* (SRG). We define SRG and show how to derive it offline from a full history of *occupancy grids*, a type of lattice-based representations. Meanwhile, we anticipate a neural network model that estimates a SRG in real time from a recent motion history of surround objects. We also describe how to use SRG in a linear programming for motion planning. We argue that SRG provides a good balance among expressivity, cost, and performance.

Lastly, to validate our method, we collected about 20 hours of aerial videos recording traffic in understructured road environments. We show how to build a validation testbed using the video dataset. For open research and development, we will publish the data and the testbed as the *Berkeley Deep Drive Drone* (B3D) dataset in summer 2020.

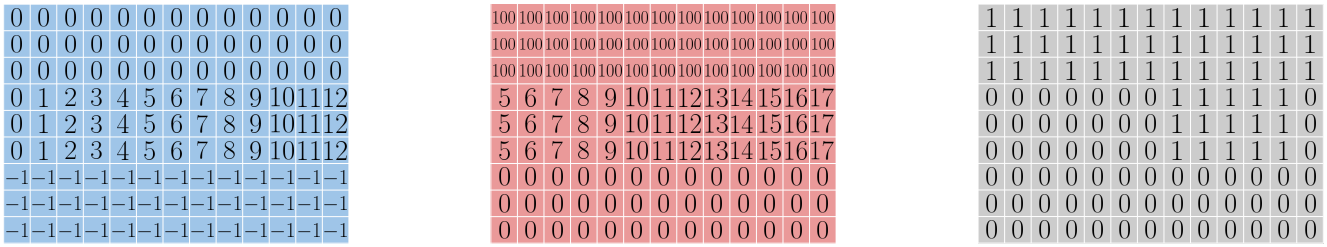
II. Proposed Approach

In this section we describe (1) what is SRG, (2) how to compute it offline from occupancy grids, (3) how to estimate it online by feeding a recent motion history of surround objects into a neural network model, (4) how to apply it in a linear program for motion planning, and (5) how does SRG balance among expressivity, cost, and performance.

*This work was supported by Berkeley Deep Drive and Berkeley Aritificial Intelligence Research.

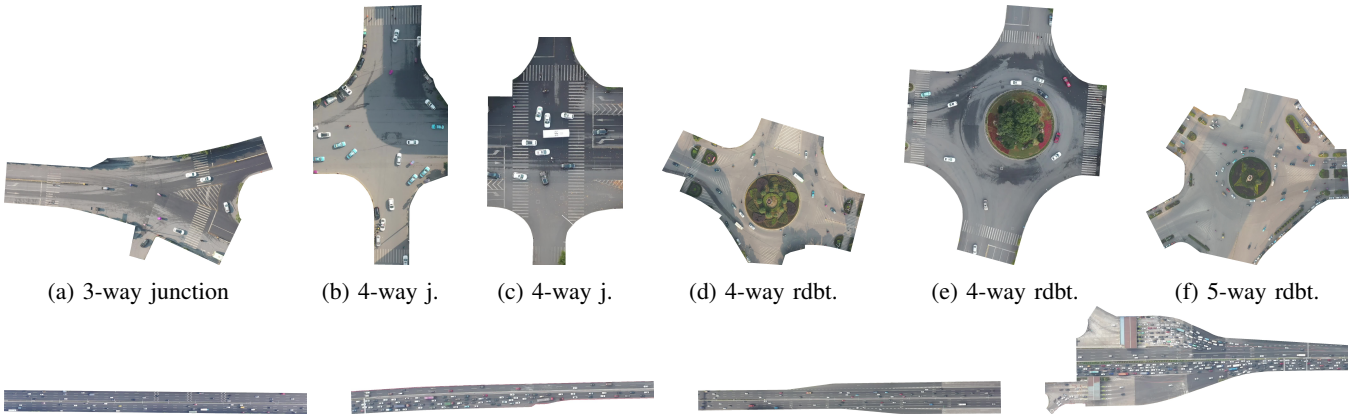
[1]The authors are with the Department of Electrical Engineering and Computer Sciences at the University of California, Berkeley.

† Fangyu Wu is the corresponding author. For questions, please contact him at fangyuwu@berkeley.edu.



(a) SRG arrival channel (b) SRG departure channel (c) Occupancy grid at $t = 9$

Fig. 1: A comparison between SRG and occupancy grid



(a) 3-way junction (b) 4-way j. (c) 4-way j. (d) 4-way rdbt. (e) 4-way rdbt. (f) 5-way rdbt.

(g) Highway (h) Highway with bottleneck (i) Highway w/ narrow ramps (j) Highway w/ wide ramps

Fig. 2: Selected scenes from the B3D dataset

A. *Definition*

To properly represent understructured road environments, one needs to (1) record sufficient features from the environments for decision making and to (2) leave out unnecessary details to ensure efficient computation. Additionally, (3) it will be beneficial if such representation and the derived planning models possess sufficient structure amenable for analysis and verification. Motivated by the three requirements, we develop a novel representation named SRG.

SRG is a data structure built upon the concept of occupancy grid. Like occupancy grid, SRG discretizes space into a lattice of elements or pixels. However, unlike occupancy grid, which designates a zero or one at each element to indicate if that element location is occupied, each element x in SRG stores a stack of reservation tuples $\{(t_a^{i,x}, t_d^{i,x}) | i \in I\}$, where i is an object ID, I is the set of IDs of all objects, $t_a^{i,x}$ is the arrival time of vehicle i upon element x, $t_d^{i,x}$ is the departure time of vehicle i from element x. To begin with, we initialize $t_a^{i,x} = t_{\min} - 1$ and $t_d^{i,x} = t_{\min}$ for all i, where $t_{\min}$ is a predefined starting time, usually chosen to be zero. For elements that are not reached by i, we store $t_a^{i,x} = t_{\min}$ and $t_d^{i,x} = t_{\max} + 1$ at that element, where $t_{\max}$ is a predefined ending time.

To illustrate the differences between SRG and occupancy grid, we draw comparison between the two representations in Figure 1. Concretely, we define $t_{\min} = 0$ and $t_{\max} = 99$. Figure 1a and Figure 1b describes a rectangular object of dimension 5×3 elements moving from left to right at 1 element per clock tick. Figure 1c corresponds to the occupany grid at $t = 9$. To the left of the object is a static obstacle and to the right is an unoccupied space.

Note that if we take only the earliest arrival time $\min_i(t_a^{i,x})$ and the latest departure time $\max_i(t_d^{i,x})$ for all i and for all x, we will effectively form a double-channel image. We call the two-channel image as compressed stacked reservation grid (CSRG). We name the channel that stores $\min_i(t_a^{i,x})$ as the arrival channel, and the channel that stores $\max_i(t_d^{i,x})$ as the departure channel. If the objects in the scene have non-overlapping paths, CSRG will encode the same spatiotemporal constraints as SRG.

To model stochasticity in the system, one can replace scalar vector $(t_a^{i,x}, t_d^{i,x})$ with $(\mathcal{N}(\bar{t}_a^{i,x}, \sigma_a^{i,x}), \mathcal{N}(\bar{t}_d^{i,x}, \sigma_d^{i,x}))$, where $\bar{t}_a^{i,x}, \bar{t}_d^{i,x}$ represent the means of normal distributions and $\sigma_a^{i,x}, \sigma_d^{i,x}$ the corresponding standard deviations. By construction, a stochastic CSRG (SCSRG) forms a four-channel image.

B. *Offline computation*

A SRG can be computed from a continuous history of occupancy grids. Given a full record of the occupancy grids of a scene from $t_{\min}$ to $t_{\max}$, if we can assume that no two objects are overlapped in occupancy grid for all time, we will be able to identify unique objects in the scenes and to track those objects from the time of its entry to its exit. Therefore, we will be able to find all the elements those objects have

Representation	Expressivity		Cost		Performance	
	Space	Time	Computation	Memory	Feasibility	Optimality
SRG	Discrete	Continuous	Medium	Medium	Yes	Yes
Image	Discrete	Discrete	Medium	Medium	No	No
Occupancy grid	Discrete	Discrete	High	High	Yes	Yes
Graph	Discrete	Discrete	High	High	Yes	Yes
Potential field	Continuous	Continuous	Low	Low	No	No

TABLE I: Qualitative comparison among major environment representations in understructured environments

ever occupied and the time interval of occupancy. The SRG is partially formed once we fill those time intervals into the tuples for the every object. If there are static obstacles in the scene, we fill the elements of obstacles in the SRG as if they were not reached. Finally, the arrival times and departure times may be further refined beyond the temporal precision of the occupancy grid data.

C. Online estimation

To estimate SRG in real time, we conjecture that it can be learnt from an offline dataset by training a neural network model through supervised learning.

To begin with, we hypothesize that the near-term SCSRG, which is a four-channel image, can be empirically inferred from road topology, nearby vehicles' shapes, positions, velocities, and accelerations, each of which can be represented by a single-channel or double-channel feature image. Stacking the feature images together, we obtain a single, six-channel image. The task of supervised learning is therefore to find an universal function approximator that maps a six-channel image to a four-channel image.

Denote a set of images of width w, height h, and channel c to be $I_{w\times h\times c}$. To conduct the supervised learning, we need to (1) compute a training dataset offline $(I^{\text{train}}_{w\times h\times 6}, I^{\text{train}}_{w\times h\times 4})$ from the collected videos, and then to (2) train an autoencoder-like neural network to map from $I_{w\times h\times 6}$ to $I_{w\times h\times 4}$ by fitting a hyperplane into $(I^{\text{train}}_{w\times h\times 6}, I^{\text{train}}_{w\times h\times 4})$ with respect to some loss function.

For an actual autonomous vehicle, the model inputs $I_{w\times h\times 6}$ will need to be measured from a combination of cameras, LIDAR sensors, GPS, and maybe a HD map.

D. Planning with SRG

Once we obtain an estimate of the SRGs of the objects surrounding a traffic participant, we can use them for pathfinding by solving optimization as simple as a linear program. The optimization will be feasible if and only if the participants' destination is reachable from its current location.

Qualitatively, provided a vehicle *and* its desired path, the linear program can be constructed as follows. The optimization variables are the arrival times and departure times of the vehicle at each element, at which it will occupy when traversing its desired path. The objective is to minimize the final departure time of the traffic participant at the exit of the path. The dynamics can be constrained by limiting the translational and rotational speeds along the path with a set of box constraints. The solution is checked for collisions using SRGs: the departure time of the agent should be earlier than arrival times of its followers and the arrival time of the agent should be later than the the departure times of its leaders, all of which can be described by a set of linear inequality constraints. Because the objective function and all the constraints are linear, the optimization problem is therefore linear and can be solved efficiently by standard linear programming techniques.

Lastly, we note that the above linear program does not check for acceleration limits because those constraints would result in a nonlinear program, which can be significantly slower to solve, even suboptimally. Ignoring acceleration constraints may be acceptable in understructured environments, because the range of speed is often narrow and hence the acceleration limits are naturally respected. Nevertheless, if one has to impose such constraints, he or she introduces a square loss term to the objective function to penalize large accelerations so that they will never exceed limits in the agent's *operational deisgn domain*. The resulting optimization problem is often referred to as a linear quadratic regulator, which is well understood and can be solved efficiently.

E. Advantages

We argue that SRG is a highly effective representation for understructured environments because it sufficiently encodes important physical constraints in the environment without losing computational efficiency and analytical interpretability. The grid discretization scheme in space enables SRG to represent arbitrary road topology and arbitrary number of traffic participants of any shapes. The tuples stored at each elements enable SRG to log significant events in time with infinite precision and horizon. It is shown that SRG can be conveniently used in a linear program for collision checking and have clear physical interpretations amenable for theoretical analysis. A qualitative comparison of the effectiveness of SRG, image, occupancy grid, graph, and potential field in *understructured environments* is tabulated in Table I. Since SRG does not assume presence of any prescribed traffic rules, it is believed to be particularly suitable for planning generic motions on crowded, understructured roads for both vehicular and non-vehicular participants.

III. Anticipated Validation

To validate the proposed approach, we build an extensive aerial video dataset using a quadcopter, which records about 20 hours of traffic in understructured road environments. Based on the dataset, we design a validation procedure to test the efficacy of SRG for motion planning. The video dataset, the corresponding post processing code, and the testing infrastructure will be open sourced as the B3D dataset in 2020.

A. Dataset construction

Quadcopter camera is a very effective data collection approach due to its fidelity, portability, and natural respect for privacy. It can simultaneously capture movements of hundreds of vehicles at high spatial (4K) and temporal resolution (30 FPS) and a fully charged drone can continuously operate for more than 20 minutes. It can be installed virtually anywhere above the roads without disrupting existing infrastructure or affecting driving behaviors. Moreover, due to its near orthographic perspective, aerial videos naturally preserve the privacy of the recorded subjects, since almost all identifying features of the recorded subjects are concealed in the process of camera projection.

Motivated by the aforementioned advantages, we deployed a quadcopter in December 2019 and have recorded about 20 hours of aerial traffic videos on understructured urban streets and highways in China. The common theme across those videos are that the road environments are vastly understructured: traffic participants do not explicitly share a set of mutually agreed traffic rules and the right of way is only negotiated through multiagent interactions.

B. Validation testbed

To validate our approach, we propose to use the dataset in the following steps: (1) to use it to compute an offline training dataset, (2) to train an online estimation model using the training dataset, (3) to apply planning using the estimation model in a linear program, (4) to compare planned trajectories with observed trajectories. If our SRG-based planning method is effective, the computed trajectories should align well with the observed trajectories.

C. Release plan

We plan to open source the data and the codes in two releases at https://fywu85.github.io/b3d/. The first release will contain about six hours of trajectories along with the corresponding videos and codes. The second release will contain the rest 14 hours of trajectories and the corresponding videos and codes. In both releases, approximately 50% of the dataset will consist of understructured junctions and the other 50% understructured highways. A selection of the covered scenarios is illustrated in Figure 2.

Each release will be subdivided into continuous video recordings of 10 to 20 minutes in length. Each recording will contain detailed trajectories for all traffic participants, where a single trajectory is expected to ship with the following attributes: (1) timestamp, (2) elapsed time, (3) rotated bounding box, (4) length, (5) width, (6) orientation (yaw), (7) position, (8) velocity, and (9) acceleration. In near future, we also have plan to provide the GPS coordinates and the underlying connectivity graph of the roads in the videos.

The data will be released under the CDLA-Sharing-1.0 license, while the code will be released under the MIT license. The first release is expected to be come in summer 2020 and the second release is planned for fall 2020. Furthermore, as we develop more and more utilities and tests over time, we will also ship those updates to the initial releases.

IV. Applications

The SRG representation augmented by the B3D dataset can provide various utilities to the planning and simulation of automated vehicles in understructured road environments.

A. Complex planning

The first and foremost application of SRG is to augment motion planning in understructured environments, where predefined traffic rules are not present, observed, or well defined. On one hand, it may be naturally estimated from raw camera and LIDAR data using existing autoencoder-like neural network models. On the other hand, the computed SRG can be combined with linear programming to produce a feasible trajectory.

B. Realistic simulation

The trained SRG model can also be used to construct high-fidelity simulations for testing the reliability of a motion planning algorithm in unstructured environments. The large collection of the different scenarios in the B3D dataset can be used to create rich simulations of unstructured environments with accurately calibrated human behaviors and process stochasticity. With it, one can quickly test the performance of candidate motion planning algorithms to a high degree of confidence without the costs and risks of physical experiments. Furthermore, the simulations can be used in reinforcement learning to develop model-free planning agents.

V. Related Works

In this section, we provide a brief survey of related work on state-of-the-art representation design and motion planning as well as a few notable aerial traffic video datasets.

A selected list of recent work on road representations, mostly image-based, is provided as follows.

- Occupancy grid prediction [11]
- Fast and furious [2]
- Chauffeurnet [6]

Meanwhile, provided below is a summary of relevant datasets that are constructed from quadcopters in the context of motion planning.

- Standford Drone dataset [12]
- highD dataset [13]
- INTERACTION dataset [14]
- pNEUMA dataset [15]

REFERENCES

[1] C. D. of Motor Vehicles, "Waymo (avt003) rtp," 2019. [Online]. Available: https://we.tl/t-9bv5Gp8iVY

[2] W. Luo, B. Yang, and R. Urtasun, "Fast and furious: Real time end-to-end 3d detection, tracking and motion forecasting with a single convolutional net," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2018, pp. 3569–3577.

[3] A. Elfes, "Occupancy grids: A probabilistic framework for robot perception and navigation." 1991.

[4] M. Pivtoraiko, R. A. Knepper, and A. Kelly, "Differentially constrained mobile robot motion planning in state lattices," *Journal of Field Robotics*, vol. 26, no. 3, pp. 308–333, 2009.

[5] O. Khatib, "Real-time obstacle avoidance for manipulators and mobile robots," in *Proceedings. 1985 IEEE International Conference on Robotics and Automation*, vol. 2. IEEE, 1985, pp. 500–505.

[6] M. Bansal, A. Krizhevsky, and A. Ogale, "Chauffeurnet: Learning to drive by imitating the best and synthesizing the worst," *arXiv preprint arXiv:1812.03079*, 2018.

[7] M. Werling, J. Ziegler, S. Kammel, and S. Thrun, "Optimal trajectory generation for dynamic street scenarios in a frenet frame," in *2010 IEEE International Conference on Robotics and Automation*. IEEE, 2010, pp. 987–993.

[8] P. E. Hart, N. J. Nilsson, and B. Raphael, "A formal basis for the heuristic determination of minimum cost paths," *IEEE transactions on Systems Science and Cybernetics*, vol. 4, no. 2, pp. 100–107, 1968.

[9] S. Karaman and E. Frazzoli, "Sampling-based algorithms for optimal motion planning," *The international journal of robotics research*, vol. 30, no. 7, pp. 846–894, 2011.

[10] J. Borenstein, Y. Koren, *et al.*, "The vector field histogram-fast obstacle avoidance for mobile robots," *IEEE transactions on robotics and automation*, vol. 7, no. 3, pp. 278–288, 1991.

[11] S. Hoermann, M. Bach, and K. Dietmayer, "Dynamic occupancy grid prediction for urban autonomous driving: A deep learning approach with fully automatic labeling," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 2056–2063.

[12] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *European conference on computer vision*. Springer, 2016, pp. 549–565.

[13] R. Krajewski, J. Bock, L. Kloeker, and L. Eckstein, "The highd dataset: A drone dataset of naturalistic vehicle trajectories on german highways for validation of highly automated driving systems," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2118–2125.

[14] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kummerle, H. Konigshof, C. Stiller, A. de La Fortelle, *et al.*, "Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps," *arXiv preprint arXiv:1910.03088*, 2019.

[15] E. Barmpounakis and N. Geroliminis, "On the new era of urban traffic monitoring with massive drone data: The pneuma large-scale field experiment," *Transportation Research Part C: Emerging Technologies*, vol. 111, pp. 50–71, 2020.